Efficient Linear-Scaling Density Functional Theory for Molecular Systems

Khaliullin, Rustam Z ; VandeVondele, Joost ; Hutter, Juerg

Abstract: Despite recent progress in linear scaling (LS) density function theory (DFT), the computational cost of the existing LS methods remains too high for a widespread adoption at present. In this work, we exploit nonorthogonal localized molecular orbitals to develop a series of LS methods for molecular systems with a low computational overhead. High efficiency of the proposed methods is achieved with a new robust two-stage variational procedure or by replacing the optimization altogether with an accurate nonself-consistent approach. We demonstrate that, even for challenging condensed-phase systems, the implemented LS methods are capable of extending the range of accurate DFT simulations to molecular systems that are an order of magnitude larger than those previously treated.

DOI: https://doi.org/10.1021/ct400595k

Efficient linear-scaling density functional theory for molecular systems

Rustam Z. Khaliullin,[1,2,*] Joost VandeVondele,[3,†] and Jürg Hutter[1,‡]

[1] *Physical Chemistry Institute, University of Zürich, Winterthurerstrasse 180, 8057 Zürich, Switzerland*

[2] *Institute of Physical Chemistry, Johannes Gutenberg University of Mainz, Staudingerweg 7, 55128 Mainz, Germany*

[3] *Department of Materials, Swiss Federal Institute of Technology, Wolfgang-Pauli-Strasse 27, 8093 Zürich, Switzerland*

(Dated: August 19, 2013)

Despite recent progress in linear scaling (LS) density function theory (DFT) the computational cost of the existing LS methods remains too high for a widespread adoption at present. In this work, we exploit nonorthogonal localized molecular orbitals to develop a series of LS methods for molecular systems with a low computational overhead. High efficiency of the proposed methods is achieved with a new robust two-stage variational procedure or by replacing the optimization altogether with an accurate non-self-consistent approach. We demonstrate that even for challenging condensed phase systems, the implemented LS methods are capable of extending the range of accurate DFT simulations to molecular systems that are an order of magnitude larger than those treated before.

I. INTRODUCTION

Intermolecular interactions determine physical and chemical properties of a broad class of important systems such as liquids, solutions and molecular solids. Because of the broad importance of molecular systems, there is a considerable interest in developing theoretical approaches for describing interactions of weakly bonded ensembles of molecules. Simulations based on density functional theory (DFT) are already playing an important role in computational studies of gas-phase molecular clusters and condensed phase systems[1]. However, the applicability of DFT to large systems is severely limited by poor scaling of the conventional diagonalization-based Kohn-Sham (KS) DFT, the computational complexity of which grows cubically with the number of molecules.

A number of alternative methods have been proposed, which explore the natural sparsity of the one-electron density matrix (DM)[2,3] and are capable of yielding linear scaling (LS) for large systems[4–11]. However, the variational optimization of the DM is very inefficient for accurate DFT calculations[3,11], which require many basis functions per atom. Therefore, the applications of DM-based LS methods have been limited to minimal-basis tight-binding problems. The optimal basis variants of the DM methods[12–15] designed to address this issue contract the large basis set into a small number of new localized basis functions and then optimize the DM in the contracted basis. Although such methods have been successfully used for the evaluation of accurate DFT energies of very large systems[16–19] their application in long simulations is hampered by the computationally costly optimization of both the contracted orbitals and the density matrix. From this point of view, methods based on the direct optimization of localized KS orbitals[20–28] are advantageous since they require only the occupied orbitals and, thus, deal with fewer variational degrees of freedom than the DM methods. Unfortunately, the progress in the development of orbital-based LS methods has been hindered by the inherently difficult convergence of the localized-orbital optimization[3,21,24,26].

Hence, the computational cost of the existing LS methods[3,19,29,30] remains too high to provide a competitive alternative to the conventional cubic scaling DFT for routine molecular dynamics (MD) or Monte Carlo (MC) simulations. In this work, we present a series of efficient LS DFT methods for molecular systems based on localized nonorthogonal MOs. To reduce the computational overhead of these methods, we developed a two-stage variational procedure that exhibits fast and stable convergence and, thus, overcomes the major obstacle to practical applications of orbital-based LS DFT. We also show that for weakly bonded molecular systems the second stage of the variational optimization can be replaced, without an appreciable loss of accuracy, with an even faster local non-self-consistent approach. The accuracy and computational performance of the proposed methods are analyzed using ice and liquid water as representatives of a broad class of condensed phase molecular systems – the most challenging application for LS methods.

II. THEORETICAL APPROACH

In the first step, the electrons of a system are logically divided into non-overlapping subsets called groups. For molecular systems, a typical group consists of all electrons of a single molecule, all atoms of which are referred to as group's centers. Within the KS DFT electrons are described by one-electron molecular orbitals (MOs) $|\psi_{xi}\rangle$, where the indices indicate that electron i belongs to group x.

In the next step, a localization domain is specified for each group of electrons. In this work, the localization domain of group x is chosen as a subset of atom-centered basis orbitals (AOs) $|\phi_{x\mu}\rangle$ of all neighboring molecules. Two molecules are considered neighbors if there is a pair of their atoms located within a sum of the element-specific cutoff radii R_c.

To describe a localization domain in the one-electron

Hilbert space it is convenient to introduce the following projection operator:

$$\hat{I}_x = \hat{I}_x^\dagger = |\phi_{x\mu}\rangle S^{x\mu,x\nu}\langle\phi_{x\nu}|, \quad (1)$$

where $S^{x\mu,x\nu}$ are matrix elements of the inverse of the AO overlap submatrix $S_{x\mu,x\nu} = \langle\phi_{x\mu}|\phi_{x\nu}\rangle$. Note that summation is implied over orbital indices but not over group or domain indices throughout the paper.

Finally, locality constraints are imposed on the occupied MOs of each group by restricting their expansion only to the AOs of their own domains. This restriction can be expressed conveniently with a domain projector:

$$|\psi_{xi}\rangle \equiv \hat{I}_x|\psi_{xi}\rangle = |\phi_{x\mu}\rangle T^{x\mu}{}_{xi}. \quad (2)$$

This leads to nonorthogonal MOs, which are localized on their domains to the same extent as AOs are localized on atoms and, thus, called absolutely localized molecular orbitals (ALMOs)[31].

By construction an AO may belong to several localization domains and, therefore, the Hilbert space spanned by AOs is partitioned into overlapping subspaces:

$$\hat{I} \neq \sum_x \hat{I}_x. \quad (3)$$

It has been shown that, in the case of such overlapping subspaces, a straightforward variational optimization of the localized orbitals given by Eq. (2) exhibits extremely slow convergence[3,21,24,26,32] rendering such an approach impractical (Figure 1).

To circumvent the convergence problem we propose a two-stage optimization procedure. In the first stage, the localization radius R_c is set to zero and the variational optimization is performed for ALMOs expanded only in terms of AOs of their centers. For this case of non-overlapping domains the optimization can be carried out efficiently and yields zero-order orbitals $|\psi^0_{xi}\rangle$. These orbitals represent each molecule's electrons polarized by the field of all other molecules in the system with the intermolecular charge transfer restricted by the $R_c = 0$ constraint.

In the second stage, R_c is set to a desired finite value and zero-order orbitals and the corresponding density operator $\hat{R}^0$ are used to construct trial orbitals of the following form:

$$\begin{aligned}|\psi_{xi}\rangle &= |\psi^0_{xi}\rangle + \hat{I}_x(\hat{I} - \hat{R}^0)\hat{I}_x\hat{X}|\psi^0_{xi}\rangle \\ &= |\psi^0_{xi}\rangle + \hat{I}_x(\hat{I} - \hat{R}^0)|\phi_{x\mu}\rangle X^{x\mu}{}_{xi}. \end{aligned} \quad (4)$$

In Eq. (4), $\hat{X}$ is a single-electron excitation operator, the matrix elements of which $X^{x\mu}{}_{xi} = \langle\phi^{x\mu}|\hat{X}|\psi^0_{xi}\rangle$ describe delocalization of zero-order orbitals into AOs of the neighbors and represent variational degrees of freedom in the second stage of the optimization. Operator $\hat{I}^x$ ensures that orbitals remain absolutely localized within their domains whereas $(\hat{I} - \hat{R}^0)$ restricts the delocalization to the zero-order unoccupied subspace. Keeping delocalization corrections orthogonal to zero-order orbitals resolves the problem of slow optimization observed for unprojected orbitals in Eq. (2) (Figure 1).

The importance of keeping the orbitals orthogonal to a non-overlapping reference has also been emphasized in Ref.[33]. The guidelines in Ref.[33] state that the non-overlapping reference regions should be chosen close to the centers of the maximally localized Wannier functions in the system[33,34]. Since these are not always known *a priori* one has to rely on chemical intuition to construct the reference. In contrast, the methods presented here can be used as a black-box procedure that generates a variationally optimal zero-order state.

The optimization of the projected trial orbitals is performed to minimize either the KS energy or a simplified energy functional with the fixed zero-order KS Hamiltonian $\hat{H}^0$:

$$E^c = \mathrm{Tr}\left[(\hat{R} - \hat{R}^0)\hat{H}^0\right] \quad (5)$$

The former approach is variational and will be referred to as ALMO SCF. The latter method designated as ALMO(X) follows the Harris approach to molecular systems[35], in which the Hamiltonian and the occupied orbitals are not updated to reach self consistency. In ALMO(X), the final energy is obtained as the corrected zero-order energy:

$$E = E^0 + \min_X E^c \quad (6)$$

Furthermore, since typical interactions between molecules are weak an accurate density $\hat{R}$ in Eq. (5) can be obtained efficiently using a non-iterative procedure. Inspired by an early work on localized MOs[31], we propose to construct $\hat{R}$ from the lowest eigenvectors of the locally-projected zero-order KS Hamiltonians, which are constructed and diagonalized for each domain:

$$\hat{H}^{\mathrm{LP}}_x = (\hat{I} - \hat{R}^0 + \hat{R}^{0\dagger}_x)\hat{H}^0(\hat{I} - \hat{R}^0 + \hat{R}^0_x), \quad (7)$$

where $\hat{R}^0_x = |\psi^{xi}_0\rangle\langle\psi^0_{xi}|$ is a non-Hermitian equivalent of the local density operator[31,36]. In this approach, further denoted as ALMO(D) the final energy is obtained as $E = E^0 + E^c$.

It is important to note that the presented logical partitioning with the subsequent construction of localized MOs is a rather general approach used in a large number of electronic structure theories, which are collectively known as fragmentation methods[37]. These methods vary greatly in how the partitioning and recombination are performed and, therefore, differ in accuracy and computational cost. We would like to emphasize that in our approach a proper quantum mechanical description of the entire system is constructed in the form of the total idempotent density matrix:

$$\hat{R} = \sum_x |\psi^{xi}\rangle\langle\psi_{xi}|. \quad (8)$$

This approach provides the most rigorous description of the electronic structure.

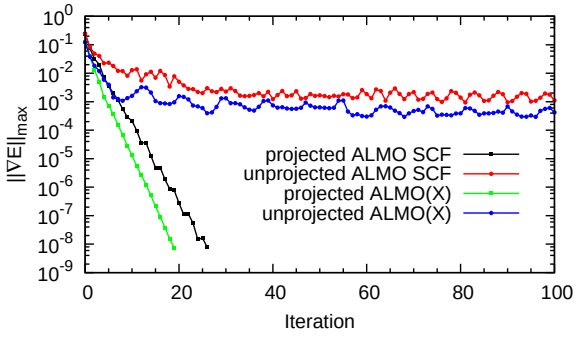

FIG. 1. **Maximum norm of the energy gradient in the second-stage PCG optimization.** The test is performed with the BLYP XC functional and TZV2P basis set for a system of 64 water molecules.

III. IMPLEMENTATION

All newly proposed methods were implemented in the CP2K package[38], which relies on the mixed Gaussian and plane wave (GPW) representation of the electronic degrees of freedom[39]. The GPW approach makes CP2K uniquely suited for orbital-based LS methods because the localized Gaussian AOs provide an accurate representation for localized MOs with just a few basis functions, whereas the use of plane waves ensures a fast LS construction of the KS Hamiltonian for large systems.

For variational optimization of zero-order orbitals, we implemented a parallel version of the previously developed DIIS-accelerated SCF MI procedure[36]. SCF MI is based on the diagonalization of the locally projected Hamiltonian submatrices, the construction of which become LS when the KS and AO overlap matrices are sparse (see Ref.[36] for details).

A preconditioned conjugate-gradient (PCG) procedure was implemented to find the optimal orbitals in the second stage of the ALMO SCF and ALMO(X) methods. The following expressions for the gradient and preconditioner

$$\frac{\partial E}{\partial X^{x\mu}{}_{xi}} = \langle \phi_{x\mu} | (\hat{I} - \hat{R}^0)\hat{I}_x(\hat{I} - \hat{R})\hat{H}_n | \psi^{xi} \rangle, \qquad (9)$$

$$P_{x\mu,x\nu} = \langle \phi_{x\mu} | (\hat{I} - \hat{R}^0)\hat{I}_x(\hat{I} + \hat{H}^0)\hat{I}_x(\hat{I} - \hat{R}^0) | \phi_{x\nu} \rangle, \qquad (10)$$

were used. Here, $\hat{H}_n$ is updated on every iteration for ALMO SCF or remains fixed to $\hat{H}^0$ for ALMO(X). As shown in Figure 1 the PCG optimization converges rapidly if the trial orbitals are represented by Eq. (4). In contrast, it is difficult to achieve convergence for the unprojected localized MOs given by Eq. (2).

An important consequence of the ALMO constraints is that both the gradient and the preconditioner are represented by the submatrices confined to their domains. The size of domains is determined only by R_c and does not change with the number of molecules. Therefore, the computational cost of the PCG optimization exhibits a linear growth in the limit of large systems.

A special care was taken to reduce the cost of evaluating matrices in Eqs. (9) and (10) for systems that are too small to exhibit sparsity. To this end, the density matrix was expressed in terms of a smaller ALMO coefficient matrix and the order of matrix multiplications was chosen to avoid steps that scale cubically with the size of the AO basis set. This approach allowed us to reduce greatly the computational overhead of the LS algorithm for large AO basis sets.

In ALMO(D), the construction of the locally-projected KS matrix can be performed as described in the previous work[36]. Although Ref.[36] deals with nonorthogonal MOs localized strictly on their molecules, it can be shown[31] that the same algorithm is applicable to the case of overlapping domains. The diagonalization of the locally-projected KS matrices is done independently for each domain maintaining the LS behavior of the method.

It is important to note that the construction of the biorthogonal occupied orbitals $|\psi^{xi}\rangle$ requires the inversion of the ALMO overlap matrix. Although the size of this matrix is small and independent of the size of the AO basis set it is not confined to individual domains. This inversion was carried out using the iterative Hotelling method[40] that is based entirely on matrix multiplications and is LS for large sparse matrices.

All parallel sparse matrix multiplications were performed with the DBCSR library implemented in CP2K and briefly described elsewhere[11].

IV. RESULTS AND DISCUSSION

Accuracy. There are several types of error introduced in the proposed ALMO methods. First, the error due the localization constraints imposed by operator $\hat{I}_x$ is present in all ALMO methods. Second, the non-self-consistent treatment of electron delocalization over neighbors introduces additional errors in ALMO(X) and ALMO(D).

To assess the accuracy of the new methods we calculated the energies for 100 decorrelated snapshots collected from a 10 ps MD simulation of liquid water performed at constant temperature (300 K) and density (0.9966 g/cm^3) with the conventional DFT. Molecular orbitals were represented by a triple-ζ Gaussian basis set with two sets of polarization functions (TZV2P) – a basis sufficient for an accurate description of liquid water[41]. A cutoff of 400 Ry was used to describe the electron density. The exchange-correlation (XC) energy was approximated with the BLYP functional[42,43]. The Brillouin zone was sampled at the Γ-point and separable norm-conserving pseudopotentials were used to describe the interactions between the valence electrons and the ionic cores[44]. The periodic simulation cell contained 64 water molecules.

Energy errors were calculated relative to the conventional KS energies $\Delta E \equiv E_{ALMO} - E_{KS}$. Their distribution for the 100 snapshots were characterized by the mean

error $\langle \Delta E \rangle$ and standard deviation $\sigma = \sqrt{\langle \Delta E - \langle \Delta E \rangle \rangle}$, which are shown in Figure 2 as a function of the localization radius R_c. Note that the element-specific cutoff radii will be further expressed in units of the elements' van der Waals radii (vdWR).

The mean error in Figure 2b demonstrates that non-self-consistent methods can give energies that are lower than the variational values. For simulations, in which coordination number of molecules does not change drastically, the mean error represents a constant shift of the potential energy surface and, on average, does not affect the behavior of molecules. In such cases, the quality of ALMO methods can judged by σ (Figure 2c).

For liquid water, restricting electrons to their own molecules (i.e. $R_c = 0$) results in the unacceptable error with $\langle \Delta E \rangle = 32.1$ and $\sigma = 1.3$ kJ/mol per molecule. However, allowing electron delocalization over just the first coordination shell ($R_c = 1.2$ vdWR, see Figure 2a) produces rather accurate results with $\sigma \sim 0.05$ kJ/mol per molecule. Electron delocalization over more distant neighbors decreases the localization error even further. For instance, if electron delocalization is allowed over the first and second coordination shells ($R_c \approx 1.6$ vdWR) the mean error per hydrogen bond drops to less than 1% of its total strength. For this cutoff radius, σ reaches the asymptotic value of ~0.02 kJ/mol for the non-SCF ALMO(X) and ALMO(D) methods. It is important to note that σ drops to zero as $R_c \to \infty$ for the self-consistent ALMO SCF method.

To estimate the effect of these errors on the structural properties of water we performed MC simulations[45] using the ALMO methods and calculated the radial distribution functions (RDF). MC simulations were used in this case because the atomic forces of the ALMO energies required for MD simulations have not been implemented yet. The accurate settings of the MD simulation described above were retained in the MC simulations. 45,000 rotational and translational MC moves were performed in each simulation. To increase the acceptance ratio each move was pre-sampled using TIP3P potential[46]. For this reason the intramolecular OH distances and HOH angle of water molecules were fixed throughout the simulation to their TIP3P values of 0.9572Å and 104.52°, respectively.

Figure 3 shows that the oxygen-oxygen RDF calculated with ALMO(D) converge rapidly to the reference curve with increase of R_c (see Figure S1 in the Supporting Information for a comparison of RDFs for all ALMO methods). For $R_c = 1.2$ vdWR, the RDF deviates only slightly from the reference in the region of large oxygen-oxygen distances. For $R_c = 1.6$ vdWR, the error of the ALMO(D) description of the RDF lies within simulation's statistical error.

Computational efficiency. The LS behavior of the ALMO methods applied to hexagonal ice can be clearly seen in Figure 4. The LS regime is achieved at as few as 1,000 water molecules with a rather accurate representation of AOs (TZV2P) and MOs ($R_c = 1.6$ vdWR).

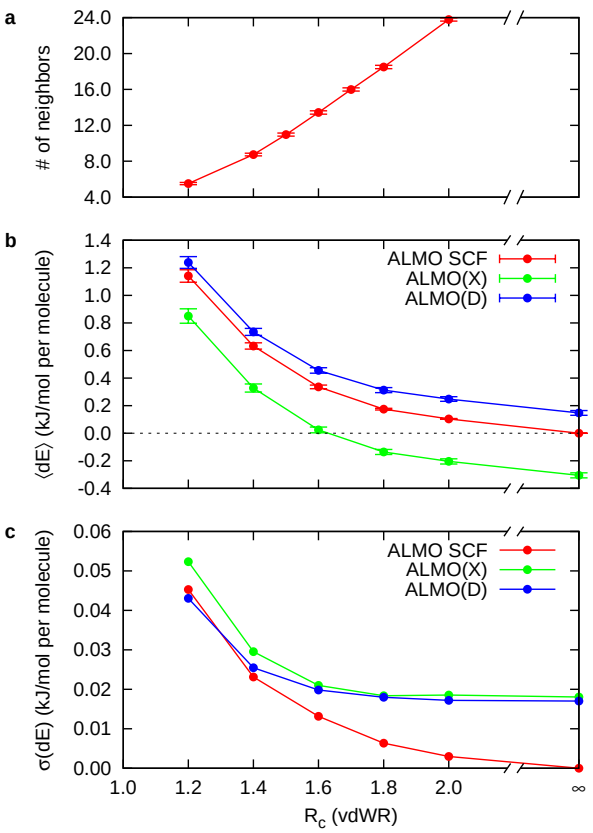

FIG. 2. **Analysis of the accuracy of the ALMO methods for liquid water.** a. Dependence of the average number of neighbors on the localization radius. b. Dependence of the mean error per molecule $\langle \Delta E \rangle$ on the localization radius. The error bars show the standard deviation. c. Dependence of the standard deviation σ per molecule on the localization radius. Calculations were performed with the BLYP XC functional and TZV2P basis set.

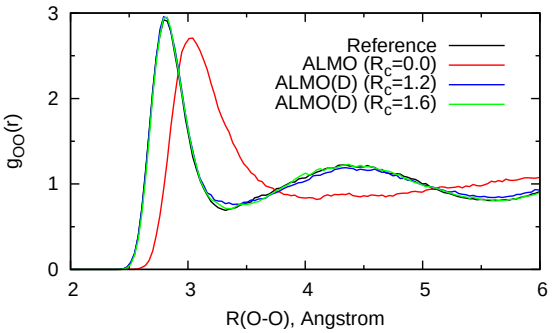

FIG. 3. **Oxygen-oxygen radial distribution functions for liquid water.** From Monte Carlo simulations performed with the BLYP XC functional and TZV2P basis set for 64 water molecules.

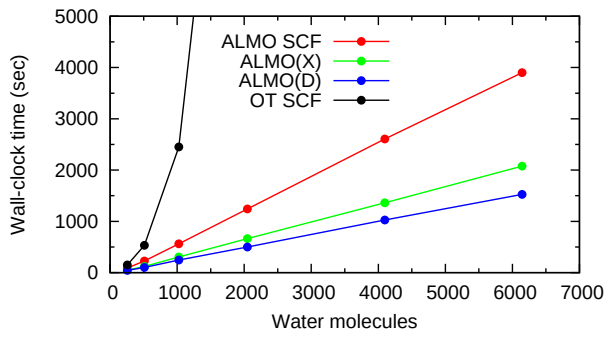

FIG. 4. **Timing benchmarks for the energy evaluation of hexagonal ice.** Energies were obtained with the BLYP XC functional and TZV2P basis set on 256 cores. The localization radius was set to 1.6 vdWR.

The non-iterative ALMO(D) method is the most efficient (i.e. has the lowest prefactor) among the newly proposed methods. The difference in performance of the iterative ALMO(X) and ALMO SCF schemes comes from the repetitive re-evaluation of the KS matrix and increased number of iterations in the latter more accurate variational method. Figure S2 in the Supporting Information shows that the LS behavior is retained for larger systems.

It is worth noting that high efficiency of the ALMO approach relies on the LS construction of the KS matrix, which is performed in the reciprocal space using plane waves[39]. Asymptotically the KS build contributes from 25 to 30% to the total time (Figure S3 in the Supporting Information).

To demonstrate the computational efficiency of the ALMO methods their performance is compared to that of the orbital transformation (OT) approach[47,48]. OT is a highly efficient and well-optimized SCF algorithm that performs PCG optimization of the occupied MOs. Matrix multiplication of the sparse Hamiltonian and AO overlap matrices with the dense representation of delocalized MOs is the most expensive computational step of OT SCF for systems considered here. Since it scales quadratically with the number of molecules the contribution of the cubic-scaling Cholesky inversion of the preconditioner increases as the size of the system grows (e.g. it reaches 23% for 4096 water molecules). This results in the intermediate quadratic-cubic scaling behavior of the OT SCF curve in Figure 5a. It should be noted that using a sparse LS Cholesky inversion does not provide any advantages in our tests because, for large basis sets like TZV2P, the inverted matrices remain dense even for systems containing many thousands of water molecules. For larger systems, OT SCF becomes inferior to a DM-based LS method discussed below. It is important to note that for comparison purpose the ALMO methods, OT SCF and the DM-based method use the same values of all common parameters (e.g. cutoff energies, grid screening thresholds, convergence targets).

Timing benchmarks for the energy evaluation of liquid water (Figure 5a) show that the ALMO methods are more efficient than OT SCF even for systems of moderate size. For systems containing 1,000 water molecules, ALMO(D) is more than an order of magnitude faster then the conventional SCF optimization (Figure 5b). Although the ALMO methods are not yet perfectly LS for 4,000 molecules (compare to the case of the lower-density hexagonal ice above) the speed-ups reach impressive two orders of magnitude for ALMO(D) and a factor of ∼25 for ALMO SCF for this system.

It is instructive to compare the computational cost of the ALMO methods to that of the LS approaches that rely on the sparsity of the DM[4–11]. As a representative of this class we used a LS method that performs a variational optimization of the DM computing it as the matrix sign function of the effective Hamiltonian[11]. Unlike the ALMO methods, this approach does not require a prescribed sparsity pattern and, therefore, reproduces the KS energies exactly. However, for extended basis sets like TZV2P, the matrices in the DM-based methods become sufficiently sparse only for large systems containing ∼10,000 water molecules (see the DM SCF curve in Figure 5a and Supporting Figure S4). For smaller systems (i.e. the dense matrix regime) any DM method is unlikely to surpass the efficiency of OT SCF let alone ALMO SCF. The reason for this is that OT SCF operates with smaller N×M matrices instead of N×N matrices, where N is basis set size and M is the number of occupied orbitals. For example, N=10×M for water and TZV2P, which explains the difference between the OT SCF and DM SCF curves in Figure 5. Given this large computational overhead we conclude that all DM-based methods are expected to be significantly slower than the approximate ALMO methods for any reasonably accurate extended basis set.

Figure 5c shows how the performance of the ALMO methods changes with R_c that controls the size of localization domains. Domain operations (e.g. the inversion of the domain preconditioners, the diagonalization of the locally-projected Hamiltonians) are currently performed with dense matrix routines, the computational cost of which grows cubically with the size of a domain. However, these routines do not significantly affect the overall performance of the ALMO methods for the physically and computationally reasonable values of R. For large molecules and large domains these routines can in principle be replaced with their LS equivalents.

Finally, we estimated the maximum system size, for which ALMO-based MD and MC simulations can be performed on modern computer platforms with routine access to $\sim 10^3$ compute cores. We used hexagonal ice as an example and assumed that the wall-clock time per simulation step should not exceed the generally accepted upper bound of 100 seconds. The wall-clock time was measured for 10 SCF iterations. This represents a typical number of iterations required to converge wavefunctions from an accurate initial guess, which is normally generated by extrapolation form the previous simulation steps. Figure 6 shows that, because of good parallel scalability

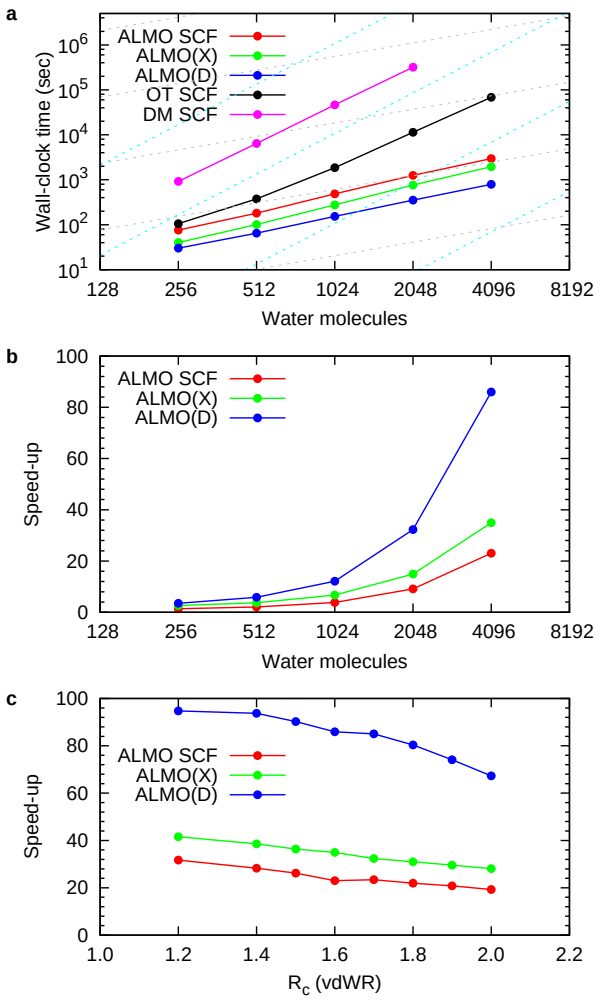

FIG. 5. **Timing benchmarks for the energy evaluation of liquid water.** Energies were obtained with the BLYP XC functional and TZV2P basis set on 256 cores. a. Wall-time required for the energy evaluation with the localization radius of 1.6 vdWR. The grey dashed lines show perfect LS whereas the cyan dashed lines mark perfect cubic scaling. DM SCF refers to a DM-based algorithm implemented in CP2K, the asymptotic LS behavior of which is better seen for the DZVP basis set calculations presented in Supporting Figure S4. b. Speed-ups for the localization radius of 1.6 vdWR relative to OT SCF. c. Speed-ups relative to OT SCF for 4096 water molecules.

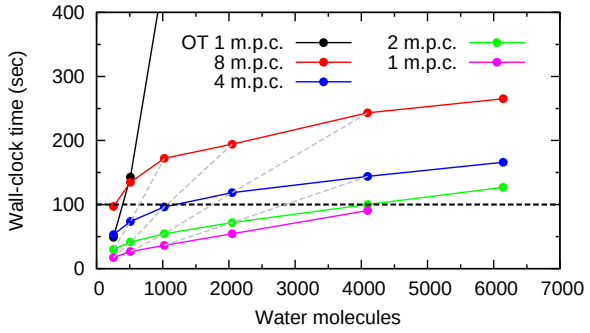

FIG. 6. **Weak scalability benchmark calculations for hexagonal ice.** The wall-clock time is measured for 10 SCF iterations for ALMO(D) and OT SCF methods using a fixed number of molecules per core (m.p.c.). Dashed grey lines connect points with the same number of cores and demonstrate linear scaling behavior on 128, 256, 512, 1024 cores. The dashed black line shows the generally accepted upper bound on the wall-clock time per simulation step. Calculations were performed with the BLYP XC functional, TZV2P basis set. The localization radius was set to 1.6 vdWR.

of the ALMO routines in CP2K the wall-clock time per simulations step can be reduced below 100 seconds for systems containing several thousands of water molecules. For comparison, the cubic scaling of OT SCF makes calculations prohibitively expensive for systems containing more than several hundred of molecules (Figure 6). This estimate roughly coincides with the size of the state-of-the-art 384-molecule simulations performed with OT SCF recently[49]. Thus, the newly proposed ALMO methods are capable of extending the range of accurate DFT simulations to molecular systems that are an order of magnitude larger than those treated with the conventional DFT methods.

It is important to mention that the wall-clock time cannot always be reduced to the 100-second mark by increasing the number of compute cores. The main reason for this is the deteriorating parallel efficiency with for an excessively large number of cores. This effect, common to all linear scaling electronic structure codes, is obvious in Figure 6: for a fixed system size, doubling the number of cores reduces does not double the efficiency of the calculations. This effect is even more pronounced for a larger number of cores, which are necessary to treat bigger systems (see Figure S5 in the Supporting Information). Despite this limitation, the low computational overhead of the developed theoretical approach enables us to reach a critical system size, from which a stable linear growth is possible in the near future when more efficient parallel libraries for matrix multiplication and/or faster intercore communication hardware become available.

V. CONCLUSIONS

We proposed a series of perturbation and variational LS methods for molecular systems based on nonorthogonal molecular orbitals with predetermined locality constraints. For the variational methods, a newly developed two-stage optimization procedure circumvents the long-standing problem of slow optimization of localized MOs and ensures fast and stable convergence of the SCF procedure. Using liquid water as an example, we showed that substantial computational savings are possible in the new approaches with a negligible loss of accuracy for energies and structural properties. Because of their low computational overhead, the proposed methods are

computationally superior to the conventional DFT algorithms even for small systems in the pre-LS regime. The computational advantage of the new methods grows with the number of molecules as they approach the asymptotic LS behavior. We demonstrated that an efficient parallel implementation of the LS algorithms offers a promising route to extend accurate DFT simulations of molecular systems to previously inaccessible size- and time-scales.

We would like to note that restricting electrons to local domains can be used not only for computational advantage but also to gain an additional physical insight into fundamental aspects of intermolecular interactions. Energy decomposition analysis and charge transfer analysis based on localized MOs[50–52] are already widely used to study the nature of bonding between gas-phase molecules. Our recent work shows that the ALMO(X) method extends the applicability of such decomposition schemes to condensed phase systems[53,54].

Although the manuscript presents applications only to condensed phase systems it is important to point out that all ALMO-based algorithms in the CP2K package are also applicable to molecular systems with the periodicity in one and two dimensions as well as to gas-phase clusters.

A number of follow-on research developments appear useful based on this work. First, the implemented localization scheme is based on the sharp cutoff radius that introduces discontinuities into potential energy surfaces. Therefore, a modification of the LS methods to smooth localization cutoffs is desirable for future MD applications. Second, the variational ALMO SCF method is in principle suitable for strongly coupled atoms. However, the currently implemented algorithms do not converge for such cases making the ALMO SCF approach impractical for covalent systems. Third, since the proposed methods are based on a division of a big system into smaller subsystems, they might be helpful for combining different electronic structure methods, such as Møller-Plesset perturbation theory and DFT, in a single system.

Acknowledgments. The research was funded by the Swiss University Conference through the High Performance and High Productivity Computing (HP2C) Programme. The authors are grateful to the Swiss National Supercomputing Centre (CSCS) for computer time. R.Z.K. would also like to thank the Swiss National Science Foundation for financial support.

Supporting Information Calculated radial distribution functions of liquid water, comparison of timing benchmarks for the DZVP and TZV2P basis sets, timing benchmarks for systems containing 32,768 water molecules, timing benchmarks for the Kohn-Sham matrix build. This material is available free of charge via the Internet at http://pubs.acs.org.

* rustam@khaliullin.com
† joost.vandevondele@mat.ethz.ch
‡ hutter@pci.uzh.ch

[1] Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; University Press: Cambridge, 2009.
[2] Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168.
[3] Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.
[4] Li, X. P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891.
[5] Lee, T. S.; York, D. M.; Yang, W. T. *J. Chem. Phys.* **1996**, *105*, 2744.
[6] Li, X. S.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J.; Schlegel, H. B. *J. Chem. Phys.* **2003**, *119*, 7651.
[7] Palser, A. H. R.; Manolopoulos, D. E. *Phys. Rev. B* **1998**, *58*, 12704.
[8] Larsen, H.; Olsen, J.; Jorgensen, P.; Helgaker, T. *J. Chem. Phys.* **2001**, *115*, 9685.
[9] Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. *J. Chem. Phys.* **2003**, *118*, 8611.
[10] Shao, Y.; Saravanan, C.; Head-Gordon, M.; White, C. A. *J. Chem. Phys.* **2003**, *118*, 6144.
[11] Vandevondele, J.; Borstnik, U.; Hutter, J. *J. Chem. Theory Comput.* **2012**, *8*, 3565.
[12] Hierse, W.; Stechel, E. B. *Phys. Rev. B* **1994**, *50*, 17811.
[13] Hernandez, E.; Gillan, M. J. *Phys. Rev. B* **1995**, *51*, 10157.
[14] Hernandez, E.; Gillan, M. J.; Goringe, C. M. *Phys. Rev. B* **1996**, *53*, 7147.
[15] Mostofi, A. A.; Haynes, P. D.; Skylaris, C.-K.; Payne, M. C. *J. Chem. Phys.* **2003**, *119*, 8842.
[16] Hine, N. D. M.; Haynes, P. D.; Mostofi, A. A.; Skylaris, C. K.; Payne, M. C. *Comput. Phys. Commun.* **2009**, *180*, 1041.
[17] Bowler, D. R.; Miyazaki, T. *J. Phys.: Condens. Matter* **2010**, *22*, 074207.
[18] Hine, N. D. M.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. *J. Chem. Phys.* **2010**, *133*, 12.
[19] Bowler, D. R.; Miyazaki, T. *Rep. Prog. Phys.* **2012**, *75*, 036503.
[20] Galli, G.; Parrinello, M. *Phys. Rev. Lett.* **1992**, *69*, 3547.
[21] Mauri, F.; Galli, G.; Car, R. *Phys. Rev. B* **1993**, *47*, 9973.
[22] Ordejon, P.; Drabold, D. A.; Grumbach, M. P.; Martin, R. M. *Phys. Rev. B* **1993**, *48*, 14646.
[23] Mauri, F.; Galli, G. *Phys. Rev. B* **1994**, *50*, 4316.
[24] Ordejon, P.; Drabold, D. A.; Martin, R. M.; Grumbac, M. P. *Phys. Rev. B* **1995**, *51*, 1456.
[25] Kim, J. N.; Mauri, F.; Galli, G. *Phys. Rev. B* **1995**, *52*, 1640.
[26] Fattebert, J. L.; Gygi, F. *Comput. Phys. Commun.* **2004**, *162*, 24.
[27] Fattebert, J. L.; Gygi, F. *Phys. Rev. B* **2006**, *73*, 115124.
[28] Burger, S. K.; Yang, W. T. *J. Phys.: Condens. Matter* **2008**, *20*, 294209.
[29] Parr, R. G.; Yang, W. T. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701.
[30] Bowler, D. R.; Miyazaki, T.; Gillan, M. J. *J. Phys.: Condens. Matter* **2002**, *14*, 2781.
[31] Stoll, H.; Wagenblast, G.; Preuss, H. *Theor. Chim. Acta* **1980**, *57*, 169.
[32] Bowler, D. R.; Gillan, M. J. *Mol. Simul.* **2000**, *25*, 239.

[33] Tsuchida, E. *J. Phys. Soc. Jpn.* **2007**, *76*, 034708.
[34] Tsuchida, E. *J. Phys.: Condens. Matter* **2008**, *20*, 294212.
[35] Harris, J. *Phys. Rev. B* **1985**, *31*, 1770.
[36] Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. *J. Chem. Phys.* **2006**, *124*, 204105.
[37] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2011**, *112*, 632.
[38] CP2K Open Source Molecular Dynamics. http://www.cp2k.org (accessed Aug 18, 2013).
[39] Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103.
[40] Hotelling, H. *Annals of Math. Stat.* **1943**, *14*, 1.
[41] Vandevondele, J.; Mohamed, F.; Krack, M.; Hutter, J.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **2005**, *122*, 014515.
[42] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
[43] Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
[44] Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641.
[45] Kuo, I. F. W.; Mundy, C. J.; Mcgrath, M. J.; Siepmann, J. I.; Vandevondele, J.; Sprik, M.; Hutter, J.; Chen, B.; Klein, M. L.; Mohamed, F.; Krack, M.; Parrinello, M. *J. Phys. Chem. B* **2004**, *108*, 12990.
[46] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
[47] Vandevondele, J.; Hutter, J. *J. Chem. Phys.* **2003**, *118*, 4365.
[48] Weber, V.; Vandevondele, J.; Hutter, J.; Niklasson, A. M. N. *J. Chem. Phys.* **2008**, *128*, 084113.
[49] Kühne, T. D.; Pascal, T. A.; Kaxiras, E.; Jung, Y. *J. Phys. Chem. Lett.* **2010**, *2*, 105.
[50] Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. *J. Phys. Chem. A* **2007**, *111*, 8753.
[51] Mo, Y. R.; Gao, J. L.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530.
[52] Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 184112.
[53] Kühne, T. D.; Khaliullin, R. Z. *Nat. Commun.* **2013**, *4*, 1450.
[54] Khaliullin, R. Z.; Kühne, T. D. *Phys. Chem. Chem. Phys.* **2013**, *accepted*, DOI:10.1039/C3CP51039E.